Transcriptional Mapping of the Coat Protein Gene of Tomato Golden Mosaic Virus

By I. T. D. Petti,† R. H. A. Coutts and K. W. Buck*

Department of Pure and Applied Biology, Imperial College of Science and Technology, London SW7 2BB, U.K.

(Accepted 15 February 1988)

SUMMARY

Polyadenylated RNA was isolated from Nicotiana benthamiana plants infected with tomato golden mosaic virus (TGMV). Northern hybridization with a strand-specific probe of cloned TGMV DNA A revealed a 0.9 kb transcript with the same orientation as virion DNA. The positions of the 5' and 3' termini of the transcript, which were mapped by nuclease protection and primer extension techniques, indicated that it corresponded to the virus coat protein mRNA and enabled the likely functional promoter and polyadenylation signals to be identified.

Members of the Geminivirus group of plant viruses are characterized by their double-icosahedral or 'geminate' particles, a genome of circular ssDNA molecules, 2.5 to 3.0 kb in size, and a single capsid polypeptide species M, 28000 to 30000 (Matthews, 1982). Tomato golden mosaic virus (TGMV) belongs to a subgroup of the geminiviruses that comprises viruses having bipartite genomes, whitefly vectors and dicotyledonous hosts, e.g. African cassava mosaic virus (ACMV) (synonym cassava latent virus) (Stanley & Gay, 1983; Stanley, 1983), bean golden mosaic virus (BGMV) (Morinaga et al., 1983; Howarth et al., 1985) and TGMV (Hamilton et al., 1983, 1984).

Six open reading frames (ORFs) have been identified in the genome of TGMV, four on DNA A, designated AR1, AL1, AL2 and AL3 and two on DNA B, designated BR1 and BL1, where R and L indicate a rightward or leftward orientation with respect to the 'common region', a sequence of about 200 bases which is almost identical in the two DNA components (Hamilton et al., 1984). There are counterparts for all these ORFs in the genomes of ACMV (Stanley & Gay, 1983) and BGMV (Howarth et al., 1985).

The only TGMV gene product to be identified so far is the coat protein, encoded by ORF AR1 (Kallender et al., 1988, preceding paper). We now report on the identification of a transcript spanning the coat protein gene and the location of its 5' and 3' termini.

Nicotiana benthamiana plants were grown and infected with TGMV as described (Hamilton et al., 1981; Stein et al., 1983). Infected leaf tissue was frozen with liquid nitrogen and ground to a fine powder with a mortar and pestle. The material (1 g) was then mixed with 3 ml of extraction buffer (100 mm-Tris–HCl, 25 mm-disodium EDTA, 1% SDS, 0.3% L-ascorbic acid, 1% 2-mercaptoethanol, pH 8.3) and extracted twice with an equal volume of phenol:chloroform (24:1 v/v) saturated with 100 mm-Tris–HCl pH 8.7) and once with an equal volume of chloroform:isoamyl alcohol (24:1 v/v). Nucleic acids, precipitated by addition of 0.1 vol. of 3 M-lithium acetate pH 6.0 and 2.5 vol. of ethanol and storage at −70 °C for 60 min, were recovered by centrifugation at 10000 g for 10 min at 4 °C, dried under vacuum and resuspended in TE buffer (10 mm-Tris–HCl, 1 mm-disodium EDTA, pH 8.0). Caesium chloride was then added to a

† Present address: Department of Plant Pathology, University of California, Berkeley, California 94720, U.S.A.

0000-8239 © 1988 SGM
Fig. 1. (a) Northern blot hybridized to a strand-specific probe made with the ssDNA of mBHA (−). Nucleic acids that hybridize have the same sense as TGMV ssDNA isolated from virions. Lane 1, TGMV ssDNA from virions (a gift from V. E. Stein); the positions of circular (c) and linear (lin) forms of the genomic ssDNA are indicated. Lane 2, poly(A)⁺ RNA from TGMV-infected N. benthamiana plants. Lane 3, DNA marker fragments [the number of nucleotides is indicated (kilobase ladder, Bethesda Research Laboratories)]. (b) Northern blot hybridized to a nick-translated mixture of pBH401 and bacteriophage φX174 replicative form DNA. Lane 1, TGMV ssDNA as in (a). Lanes 2 and 3, DNA from clones mBHA (−) and mTAX (−) respectively protected from S1 nuclease by hybridization with RNA from TGMV-infected plants. Lane 4, DNA marker fragments with the approximate numbers of nucleotides indicated (φX174 RF DNA HaeIII digest, New England Biolabs).

final density of 1·69 g/ml, followed by 2-mercaptoethanol to a final concentration of 1% and the solution was centrifuged at 35000 r.p.m. for 18 h in a Beckman type 65 rotor at 20 °C. The RNA pellet was resuspended in TE buffer and the RNA was precipitated by addition of 2·5 vol. of ethanol and storage at −70 °C. The precipitate was resuspended in distilled water and the poly(A)⁺ RNA was isolated by affinity chromatography on poly(U)-Sepharose (Jost & Pehling, 1976). Any remaining traces of DNA were removed by incubation with RNase-free DNase (Tullis & Rubin, 1980).

Recombinant clones in M13 vectors (Messing, 1983), containing inserts of TGMV DNA in the same (+) or opposite (−) sense as TGMV virion DNA, to be used for transcriptional mapping, were constructed using the following restriction endonuclease sites (cleavage occurs on the 3’ side of the nucleotide, numbered according to Hamilton et al., 1984): AsuII (1037); BamHI (1356); EcoRI (2251); XhoI (400) (Fig. 3). mBHA (+) and mBHA (−), which were constructed by recloning the TGMV DNA A insert of pBH401 (Bisaro et al., 1982) into the EcoRI site of M13mp7, were gifts from W. D. O. Hamilton. mTAX (+) and mTAX (−) were constructed by excising the TGMV DNA A insert of pBH401 with EcoRI, recircularizing the DNA and restricting with XhoI as described by Hayes et al. (1988) and recloning into the SalI site of M13mp8. mTAEX (+) was constructed by cloning the 320 (approx.) bp AsuII–BamHI fragment of TGMV DNA A (from mTAX (−) replicative form DNA) between the AccI and BamHI sites of M13mp19.
Short communication 1361

EcoRI/BamHI fragment of mTAX (−) replicative form DNA, containing the 737 bp EcoRI/XhoI fragment of TGMV DNA A, between the EcoRI and BamHI sites of M13mp18. 32P-labelled strand-specific probes were synthesized from recombinant M13 ssDNA templates as described by Hu &Messing (1982). DNA sequencing ladders for use as size markers were generated from recombinant M13 clones by the dideoxynucleotide procedure of Sanger et al. (1977) in conjunction with urea-polyacrylamide gel electrophoresis (Sanger & Coulson, 1978). Other DNA manipulations were as described by Maniatis et al. (1982).

Poly(A)+ RNA from TGMV-infected plants was electrophoresed in 1.5% agarose gels in BE buffer (40 mM-boric acid, 1 mM-disodium EDTA, adjusted to pH 7.6 with sodium hydroxide) containing 5 mM-methylmercuric hydroxide as a denaturant (Bailey & Davidson, 1976) and transferred to nitrocellulose membranes as described by Gal et al. (1983). To detect transcripts in the same orientation as virion DNA, the blot was hybridized (Thomas, 1980) with a 32P-labelled strand-specific probe derived from mBHA (−) ssDNA. The autoradiograph (Fig. 1a) showed a single band with an approximate size of 0.9 kb.

To find the approximate location of the 0.9 kb transcript, RNA from TGMV-infected plants was hybridized with ssDNA of mBHA (−) or mTAX (−) at 50 °C for 4 h, as described by Favalaro et al. (1980), digested with S1 nuclease and incubated with alkali to remove the RNA strand of the RNA : DNA hybrid, as described by Townsend et al. (1985). The size of each of the protected DNA fragments was determined from plots of reciprocal Mr against mobility, using denatured HaeIII restriction fragments of bacteriophage φX174 replicative form DNA as size standards and by a local 2 × 3 point method (Sealey & Southern, 1982). The size of the protected mBHA (−) fragment (Fig. 1b, lane 2) was calculated to be 785 bases. The size of the protected mTAX (−) fragment was 705 bases, indicating that the 0.9 kb transcript spanned the XhoI site of DNA A (Fig. 3). The difference in size of the two protected fragments indicates that the 5′ terminus of the transcript is about 80 nucleotides upstream of the XhoI site (nucleotide 400) i.e. at about nucleotide 320 in the numbering system of Hamilton et al. (1984).

The position of the 5′ terminus of the 0.9 kb transcript was determined more precisely by the modified nuclease protection technique of Ciliberto et al. (1983). A 32P-labelled mTAEX (+) cDNA probe was hybridized with RNA from TGMV-infected plants as above and then digested with mung bean nuclease (1000 units/ml) in 50 mM-sodium acetate pH 5.0 containing 30 mM-sodium chloride, 1 mM-zinc sulphate and 20 μg/ml sheared and denatured salmon sperm DNA (Maniatis et al., 1982) at 37 °C. Aliquots were removed after 15, 30, 45 and 60 min and extracted with phenol : chloroform : isooamyl alcohol (25 : 24 : 1 v/v/v). The RNA : DNA hybrids were precipitated by the addition of 0.1 vol. of 3 M-sodium acetate pH 5.2, 1 vol. of propan-2-ol and Escherichia coli tRNA (Sigma) to a final concentration of 10 μg/ml, recovered by centrifugation in a microfuge, dried under vacuum and resuspended in sterile distilled water. The size of the protected DNA was determined by electrophoresis in urea-polyacrylamide sequencing gels alongside sequencing ladders generated with mTAEX (+) and mTABA (+) templates. Bands of 89 and 86 nucleotides were detected which remained prominent throughout the 60 min time course of nuclease digestion (Fig. 2a). The bands were not detected in a similar experiment in which RNA from uninfected N. benthamiana plants was used to protect the probe. The 5′ termini of the protected fragments are defined at nucleotide 404 of the TGMV DNA A sequence by the XhoI/SalI junction of mTAEX. Therefore the 3′ termini of the two prominent protected fragments are at nucleotides 316 and 319 respectively.

The 5′ terminus of the 0.9 kb transcript was also located by primer extension. To obtain a suitable primer, plasmid pH404 (Bisaro et al., 1982) was digested with NcoI, the enzyme was removed by phenol extraction and the DNA was precipitated with ethanol. Restriction fragments were labelled at their 3′ ends by incubation with [32P]dCTP and DNA polymerase I Klenow fragment in 100 mM-Tris–HCl pH 7.4, 10 mM-magnesium chloride, 20 mM-mercaptoethanol at 37 °C for 30 min. The end-labelled DNA was then digested with XhoI and the 50 (approx.) bp XhoI/NcoI fragment was isolated by electrophoresis through a 12% polyacrylamide gel and electro-elution (Maniatis et al., 1982). The end-labelled XhoI/NcoI
fragment was mixed with poly(A)$^+$ RNA isolated from TGMV-infected *N. benthamiana* leaves and precipitated with ethanol. The nucleic acid was recovered by centrifugation and resuspended in hybridization buffer (Favaloro *et al.*, 1980). The solution was incubated at 85 °C for 15 min to denature the ds primer DNA fragment and then DNA:RNA hybridization was allowed to proceed at 54 °C for 4 h. The nucleic acids were precipitated twice from 70% ethanol, resuspended in 50 mM-Tris–HCl pH 7.5, 75 mM-potassium chloride, 3 mM-magnesium chloride, 10 mM-dithiothreitol, 100 μg/ml bovine serum albumin (nuclease-free, Amersham), 50 μg/ml actinomycin D, 30 units of human placental ribonuclease inhibitor (Anglian) and 200 units of Moloney murine leukaemia virus reverse transcriptase (Bethesda Research Laboratories) and the solution was incubated at 37 °C for 60 min. The reaction was terminated by extraction with phenol, the products were precipitated from 70% ethanol and their sizes were determined by electrophoresis in urea–polyacrylamide sequencing gels alongside sequencing ladders generated with mTAEX (+) and mTABA (+) templates. The primer extension products comprised a

Fig. 2. Autoradiographs of sequencing gels (a). cDNA synthesized on a mTAEX (+) ssDNA template, hybridized with RNA from TGMV-infected plants and digested with mung bean nuclease for 15 min (lane 1), 30 min (lane 2), 45 min (lane 3) and 60 min (lane 4). (b) An end-labelled 56 base primer was hybridized with poly(A)$^+$ RNA from TGMV-infected plants and extended using reverse transcriptase. Lane 1, unextended primer. Lane 2, products of the extension reaction. The lengths of size markers (in numbers of nucleotides) from sequencing ladders are indicated.
major doublet of 88 and 87 nucleotides and a less prominent triplet of 84, 83 and 82 nucleotides (Fig. 2b). The longest extension product corresponded to an RNA 5' terminus at nucleotide 317 in TGMV DNA A. It is not clear whether the different lengths are due to 5' sequence heterogeneity in the RNA or to premature termination of the reverse transcriptase by strong stop signals in the RNA sequence.

The nuclease protection and primer extension methods indicate that the 5' terminus of the 0.9 kb transcript is close to nucleotides 316 and 317. The methods are generally considered to be accurate to within two to four nucleotides (Kozak, 1984). The nucleotide sequence of this region of TGMV DNA A is 312 GATCTTTAA 320. The transcriptional start site for eukaryotic mRNA is almost invariably a purine residue with a strong preference for A; C and U are rarely if ever found (Kozak, 1984). The most likely transcriptional start site for the 0.9 kb transcript is therefore at an A residue at nucleotide 313 or 319. It is noteworthy that the smaller of the two DNA fragments protected from mung bean nuclease corresponded to a 5' terminus at nucleotide 319. Confirmation of the exact transcriptional start site would require direct sequencing of the RNA.

Analysis of the TGMV DNA A sequence (Hamilton et al., 1984) revealed potential promoter regions, corresponding to the consensus sequence TATAT/AA (Breathnach & Chambon, 1981), starting at nucleotides 89 and 287/289 (overlapping). Since TATA boxes are generally located 25 to 30 nucleotides upstream of the transcriptional initiation site in eukaryotic mRNAs (Nevins, 1983), the position of the 5' terminus of the 0.9 kb transcript at nucleotide 313 or 319 is consistent with the promoter at 287/289. In this context, it is noteworthy that a PvuI/XholI fragment containing the TATA box at 287/289, but not the one at 89, allowed expression of a neomycin phosphotransferase gene in E. coli (Petty et al., 1986).

From the most accurate determination of the length of the mBHA (−) DNA fragment protected by the 0.9 kb transcript (785 nucleotides) and the position of its 5' terminus, together with the size of the protected mTAX (−) DNA fragment (705 bases, XhoI site at nucleotide 400 in the TGMV DNA A sequence), the 3' end of the transcript [minus the poly(A) tail] maps in the region of nucleotides 1100 to 1105. Hence the transcript traverses the whole of the coat protein gene (nucleotides 327 to 1070) and is polyadenylated about 30 nucleotides beyond the termination codon (Fig. 3). Two possible transcriptional polyadenylation signals, conforming to the consensus sequence AATAAA (Proudfoot & Brownlee, 1976) were identified in the TGMV DNA A sequence starting at nucleotides 1069 and 2120 (Hamilton et al., 1984). In plants, the AATAAAA box is located 27 ± 9 bases from the poly(A) site (Joshi, 1987). Hence the former signal, which overlaps the coat protein gene termination codon, is the one more likely to be functional in vivo.

The 5' untranslated sequences of eukaryotic mRNAs range in length from three to >800 nucleotides with the majority in the range 20 to 100 nucleotides (Kozak, 1983, 1984, 1987). The leader sequence of the TGMV coat protein mRNA (eight or 14 nucleotides based on a 5' terminus at nucleotide 319 or 313) is therefore shorter than average.
Mapping of the coat protein mRNA of ACMV, a related bipartite geminivirus, indicated a transcriptional start site about 166 nucleotides upstream of the first AUG (Townsend et al., 1985). Hence ACMV coat protein mRNA has a considerably longer leader sequence than that of the TGMV coat protein mRNA. This difference may be related to a significant difference in the genome organization of the two viruses in this region. Whereas the TGMV coat protein gene starts 127 nucleotides downstream of the 200 base ‘common region’, the ACMV coat protein gene starts 246 nucleotides downstream of the corresponding region. Furthermore there is an ORF in ACMV which could encode a protein of Mr 13100, starting at nucleotide 286 (just downstream of the coat protein mRNA cap site estimated to be at about nucleotide 280) and overlapping the start of the coat protein gene in a different reading frame. The significance of this smaller overlapping ORF, which is not found in TGMV, is not known.

We thank the Science and Engineering Research Council for the award of a research grant to K.W.B. and R.H.A.C. and a studentship to I.T.D.P. Work on TGMV was done under licence no. PHF 29B/75 (106) issued by the Ministry of Agriculture, Fisheries and Food.

REFERENCES


(Received 8 January 1988)